Titanic

Felipe Smolski

26 de setembro de 2018

**Software R: capacitação em análise estatística de dados utilizando um software livre**

**Módulo Avançado**

Nome:

## Exercício Titanic - Regressão Logística

**1.** O Titanic foi um famoso navio britânico construído a partir de março de 1909 e lançado ao mar em maio de 1911. Em sua viagem inaugural em 10 de abril de 1912, cujo objetivo era partir de Southampton para Nova Iorque passando pela França e Irlanda, colidiu com um iceberg às 23h40min do dia 14 de abril. Baixe os dados do acidente em <https://vincentarelbundock.github.io/Rdatasets/csv/COUNT/titanic.csv>. Esta base de dados possui as seguintes informações:

**survived**: informa se o tripulante sobreviveu ou não (“yes”, “no”);

**class**: representa a classe em que viajavam os tripulantes (“1st class”, “2nd class” e “3rd class”).

**age**: variável que separa entre as crianças (“child”) e adultos (“adults”)

**sex**: fator com o sexo do tripulante (“women”, “man”)

**1.1.** Importe os dados para um objeto denominado “*titanic*” (lembre-se de determinar que as variávei sejam fatores).

**1.2** Crie uma nova a variável “*sobreviveu*” a partir de “*survived*” com fatores binários 0 e 1 (1 para os sobreviventes).

**1.3** Crie um modelo de regressão logística para descobrir a probabilidade de sobrevivência dos tripulantes (variável dependente sobreviveu) em relação às variáveis class, age, sex. Utilize o comando summary com o modelo. Declare a significância estatística das variáveis.

**1.4** Com base na resposta anterior, disserte sobre a direção do sinal do *log* da probabilidade de sobrevivência com relação à cada variável independente do modelo. Quem estava na primeira classe tinha maiores chances de sobrevivência, ou na terceira? Adultos ou crianças tinham melhores chances de sobreviver, homens ou mulheres?

**1.5** Transforme os coeficientes encontrados em Razão de Chances (OR - Odds Ratio), bem como verifique os intervalos de confiança.

**1.6** Com base na resposta anterior, disserte sobre a razão das chances (OR) de cada variável independente do modelo.

**1.7** Efetue a predição da probabilidade de sobrevivência de uma tripulante do sexo feminino, adulta e que viajava na primeira classe do navio.

**1.8** Efetue a predição da probabilidade de sobrevivência de um tripulante do sexo masculino, adulto e que viajava na terceira classe do navio.

**1.9** Crie a matriz de confusão e a Curva ROC para o modelo, avaliando seus dados.

**1.10** Efetue o teste de Hosmer e Lemeschow para o modelo.

**1.11** Encontre as medidas de pseudo R.

**1.12** Realize o teste de multicolinearidade para o modelo proposto. Existem indícios de multicolinearidade no modelo?

**1.13** Utilize o método **Stepwise** (*direction=‘both’*) no modelo completo e avalie se as variáveis devem permanecer no modelo ou alguma deve sair.

Envie os resultados em <https://smolski.github.io/softwarelivrer/atividades.html>.